Predictive Models for Characterization of Ecological Data

Can Ozan Tan[1*], Uygar Özesmi[2] and Bahtiyar Kurt[3]

[1]Canakkale Onsekiz Mart University, Department of Biological Sciences, 17020, Canakkale, Turkey

[2]Erciyes University, Dept. of Environmental Engineering, 38039 Kayseri, Turkey

[3]Doga Dernegi P.O. 640, Yenisehir, Ankara TR-06445, Turkey

[*]Corresponding author: Boston University, Department of Cognitive and Neural Sciences, 677 Beacon St. #201, Boston, MA 02215. Tel: +1 617 353-6741, Fax: +1 617 353 7755 e-mail: tanc@cns.bu.edu

Abstract

Although ARTMAP and ART-based models were introduced in early 70's they were not used in characterizing and classifying ecological observations. ART-based models have been extensively used for classification models based on satellite imagery. This report, to our knowledge, is the first application of ART-based methods and specifically ARTMAP for predicting habitat selection and spatial distribution of species. We compare the performance of ARTMAP to assess the breeding success of three bird species (*Lanius senator*, *Hippolais pallida*, and *Calandrella brachydactyla*) based on multi-spectral satellite imagery and environmental variables. ARTMAP is superior both in terms of performance (percent correctly classified - $pcc = 1.00$) and generalizability ($pcc > 0.96$) to those of feedforward multilayer backpropogation (> 0.87, > 0.65), linear and quadratic discriminant analysis (> 0.48, > 0.46) and k-nearest neighbor (> 0.82, > 0.66) methods. Compared to other methods, ARTMAP is able to incorporate new observations with far less computational effort and can easily add data to already trained models.
Keywords: ART; ARTMAP; artificial neural networks; backpropogation; pattern recognition; spatial habitat selection; *Lanius senator; Hippolais pallida; Calandrella brachydactyla.*

1 Introduction

Characterization of observations to explain interactions in an ecosystems as well as within communities and individual species, in order to predict a state has been one of the main problems in ecology. The inherent complexity of the ecological processes, the relatively limited number of possible observations and their susceptibility to observational and/or measurement noise has been considered among the major difficulties in predicting a state in ecology (Fielding, 1999). Subject to these constraints, efforts to characterize ecological data and predict the state of a given ecosystem or community shifted towards statistical methods, rather than box-and-arrow type differential equation models (Ross, 1976; Lassiter and Kearns, 1977). Statistical models proved to be more robust in terms of capturing nonlinearities and being generalizable over new data sets (Moilanen, 1999; DeValpine, 2003).

Several statistical techniques are readily available for the use of ecologists to characterize observations. These range from simple regression models (Gutierrez et al., 2005; Miller, 2005) to generalized additive (Dunk et al., 2004) and linear models (Özesmi and Mitsch, 1997; Tan and Beklioglu, 2005) and from classification algorithms such as k-nearest neighbor (k-NN), linear and quadratic discriminant analysis (LDA and QDA, respectively) (Joy and Death, 2003; Maron and Lill, 2004) to recently genetic algorithms (Underwood et al., 2004), pattern recognition methods, such as artificial neural networks (Lek et al., 1996; Recknagel et al., 1997; Lek and Guegan, 1999; Özesmi and Özesmi, 1999) and lately ecological data mining (Chawla et al., 2001). While standard parametric methods such as LDA, QDA and regression are mostly criticized as being dependent on strong assumptions about the distribution of the underlying data (Hastie et al., 2001), classification and pattern recognition methods require large number of training points. On the other hand, artificial neural network-based approaches are blamed to be black-box models thus not being able to provide insight into the complex interactions of the ecosystem processes, although they are able to overcome the difficulties associated with traditional statistical models (Bishop, 1995; Ripley, 1996; Hastie et al., 2001). Nevertheless, artificial neural network-based models can provide valuable insight into ecosystem dynamics as there are several techniques for ’opening the black-box’ (Özesmi and Özesmi, 1999; Olden and Jackson, 2000; Özesmi et al., 2005).

Recently, backpropogation based methods became popular in ecological applications. Their use range from characterization of habitat selection of phytoplankton (Scardi, 1996, 2001) to fish (Reyjol et al., 2001) and bird species (Özesmi and Özesmi, 1999), to modeling whole communities and ecosystems (Tan and Smeins, 1996; Tan and Beklioglu, 2005) and characterization of wildlife damage (Spitz and Lek, 1999) to gain insight into the dynamical structure of the ecosystems. However, the main drawback of backpropogation based methods has been that they are inherently off-line, that is iterative, methods using all the available data at once. In other words, each time a new observation is made, these models require to be retrained with the whole data set in order to include the new observation, thus requiring a significant amount of computational resources and time. In addition, the fact that the performance, particularly generalizability, of these methods reduces significantly with limited number of data points renders this approach to be impractical, at least in ecology where the number of observations are commonly limited.

This report aims to introduce another statistical pattern recognition model, ARTMAP, based on adaptive resonance theory (ART) (Grossberg, 1976a,b), which is relatively unfamiliar to the ecological community. ART is originally developed to explain cortico-cortical interactions for object recognition and learning in the brain during early 70’s (Grossberg, 1976a). During 80’s and early 90’s, ART was extended as a pattern recognition and classification

algorithm, and successfully applied to several benchmark technological data sets and classification of satellite imagery data (Grossberg, 1988; Carpenter et al., 1991c, 1997). However, despite its long history as a statistical pattern recognition and classification algorithm, this report, to our knowledge, is the first application of an ART based algorithm to an ecological data set. In addition to being on-line (that is a non-iterative learning algorithm, which enables easy and fast incorporation of new observations to an already trained model), ARTMAP also performs significantly better on the data set considered here, utilizing a considerably smaller amount of computational time. To that end, we used satellite-based multi-spectral data and environmental variables to predict the occurrence of three bird species of Southeastern Anatolia, namely woodchat shrike *Lanius senator* (Linnaeus, 1758), olivaceous warbler *Hippolais pallida* (Ehrenberg, 1833), and short-toed lark *Calandrella brachydactyla* (Leisler, 1814). To predict the occurrence of the three bird species we used k-NN, LDA, QDA, feedforward multilayer backpropogation network, and ARTMAP. We provide a discussion of comparative performances of these different models.

2 Methods

2.1 Traditional Classification Methods

We compared the performance of fuzzy ARTMAP model against traditional classification and pattern recognition methods commonly employed in ecological studies. The first method was k-nearest neighbor method, which is an accepted benchmark classification method, if one considers only the training data. Nearest neighbor methods use those observations in the training set $\mathcal{T}$ closest in the input space to x to form $\hat{Y}$. More specifically,

$$\hat{Y} = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \tag{2.1}$$

where $N_k(x)$ is the neighborhood of x defined by the k closest points x_i in the training sample. It is clear that when the neighborhood k is considered to be $k = 1$, k-NN methods potentially can reach the minimum classification error possible on the training set. Note that in this case the error on independent test set is intuitively expected to be quite high. In addition, we also used LDA and QDA, which are mostly argued to be "amazingly robust" on industrial data sets (Hastie et al., 2001). LDA and QDA techniques enable one to infer the posterior probabilities of the output categories based on the data observed, using Bayes theorem:

$$\mathrm{P}(G = k | X = x) = \frac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l} \tag{2.2}$$

where $f_k(x)$ is the class-conditional density of X in class $G = k$, and π_k is the prior probability of class k with $\sum_{k=1}^{K} \pi_k = 1$. LDA and QDA assume Gaussian distribution for class densities. Fundamentally, for two category cases (as in our case), and assuming that the covariances Σ_k of the class densities are equal, linear discriminant

function is given as

$$\delta_K = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k \tag{2.3}$$

where the parameters of the Gaussian distributions are estimated from the data as

$$\hat{\pi}_k = \frac{N_k}{N} \tag{2.4}$$

$$\hat{\mu}_k = \frac{\sum_{g_i=k} x_i}{N_k} \tag{2.5}$$

$$\hat{\Sigma} = \frac{\sum_{k=1}^{K} \sum_{g_i=k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T}{(N-K)} \tag{2.6}$$

where N_k is the number of class-k observations. An equivalent decision rule is given as $G(x) = \arg\max_k \delta_k(x)$. If the equality assumption of class covariances Σ_k does not hold, we obtain quadratic discriminant function

$$\delta_k(x) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) + \log \pi_k \tag{2.7}$$

with an equivalent decision boundary between each pairs of classes k and l described by a quadratic equation $\{x : \delta_k(x) = \delta_l(x)\}$. A more in-depth discussion of these two methods, among with k-NN method, can be found in Hastie et al. (2001).

Traditional classification methods has been often criticized as they require strong assumptions about the underlying distribution of the observations (Ripley, 1996; Hastie et al., 2001). To overcome this problem, connectionist artificial neural network based approaches, such as feedforward multilayer backpropogation network has become recently popular among ecological modeling (Scardi, 1996, 2001; Tan and Beklioglu, 2005). Although ART and ARTMAP family of models are another type of artificial neural networks, they differ from connectionist approaches in several aspects (Carpenter et al., 1991a,b,c, 1992). For that reason, we also compared the performance of fuzzy ARTMAP model to that of a generalized linear model (GLM) and of a multilayer feedforward backpropogation model.

2.2 ARTMAP

Briefly, ARTMAP architecture consists of two ART modules, which are self-organizing maps (Carpenter et al., 1991a), one for input space and one for output space (figure 1; ART$_a$ and ART$_b$, respectively). Learning occurs for each ART module independently, whenever an expected category matches to presented input pattern, or a novel input pattern is encountered, then categories are formed in both ART modules and mapped on an associative learning map field. Thus, ARTMAP models represent a "pseudo-supervised" learning method (Carpenter et al., 1991a). There are several variants of ART modules (Carpenter and Grossberg, 1990; Carpenter et al., 1991b,c). Here, we used fuzzy ART modules, which were developed as pattern recognition methods for data sets with continuous input space (Carpenter et al., 1991c, 1992). Shortly, each fuzzy ART system contains an input field F_0, a F_1 field receiving

bottom-up signals from F_0 and top-down input from F_2, the latter of which represents the active category (figure 1). So-called complement coding (Carpenter et al., 1992) should be employed before feeding the input vectors to fuzzy ART modules. Theoretical considerations for this requirement are discussed in detail in Carpenter et al. (1992). Fundamentally by complement coding, it is meant that an $N \times P$-dimensional input matrix $\mathbf{a}$ is coded and fed to the model as an $N \times 2P$-dimensional matrix $[\mathbf{a}, \mathbf{a}^c]$, where $a_i^c = (1 - a_i)$.

At each F_2 category node, there is a weight associated with that node, which are initially set to 1. Each weight w_{ji} is monotonically increasing with time and hence its convergence to a limit is guaranteed (Carpenter et al., 1991c, 1992). Fuzzy ART dynamics depend on a choice parameter $\alpha > 0$, a learning rate $\beta \in [0, 1]$, and a vigilance parameter $\rho \in [0, 1]$. For each given input pattern and jth node of F_2 layer, the choice function T_j is defined by

$$T_J(\mathbf{I}) = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|} \tag{2.8}$$

where $\wedge$ is the fuzzy AND operator and is equivalent to component-wise $\min$ operator, $|\cdot|$ is the euclidean norm, and $\mathbf{w}_j = (w_{j1} \cdots w_{jM})$. The system makes a category choice when at most one F_2 node can become at a given time, and the category choice is given as $T_J = \max\{T_j : j = 1 \ldots N\}$. In a choice system, the activity of a given node at F_1 layer is given as $\mathbf{x} = \mathbf{I}$ if F_2 node is inactive and $\mathbf{x} = \mathbf{I} \wedge \mathbf{w}_J$ if Jth F_2 node is selected. Resonance occurs in the ART module if

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho \tag{2.9}$$

and reset occurs otherwise. If reset occurs, the value of the choice function T_J is set to 0, and a new index J is chosen. The search process continues until the chosen J satisfies the resonance criterion (equation 2.9). Once search ends and resonance occurs, the weight vector $\mathbf{w}_J$ is updated by

$$\mathbf{w}_J^{(\text{new})} = \beta \left(\mathbf{I} \wedge \mathbf{w}_J^{(\text{old})} \right) + (1 - \beta) \mathbf{w}_J^{(\text{old})} \tag{2.10}$$

As briefly mentioned above, fuzzy ARTMAP model consists of two fuzzy ART modules, one for input and one for target vectors linked by an associative learning network and an internal controller. With reference to figure 1, when a prediction by ART$_a$ module, which receives the input vectors, is disconfirmed at ART$_b$ module, receiving target vector, inhibition of map field activation induces the match tracking process, which raises the ART$_a$ vigilance ρ_a to just above the F_1^a activation so that the activation of F_0^a matches the reset criterion (i.e., ρ_a is decreased just to miss the match criterion given by equation 2.9). This triggers an ART$_a$ search process which leads to activation of either an ART$_a$ category that correctly predicts b at match field, or to activation of a new node which has not used before (that is, either an already formed category that predicts b is selected, or a new category is created). ART and ARTMAP algorithms, in essence, are similar to k-NN methods with adaptive update of the size of the neighborhood with each pattern encountered in the data. It is, nevertheless, a nonlinear algorithm such that the shape of the clusters built based on the patterns embedded in the input space are nonlinear. For details of fuzzy ART algorithm as well as for its geometrical interpretation, readers are referred to Carpenter et al. (1991c), and the details of fuzzy ARTMAP algorithm can be found in Carpenter et al. (1992).

Although new to ecology, ART and ARTMAP theory has been developed since early 70's, and the reader is referred to Cohen and Grossberg (1983) and Grossberg (1988) for theoretical considerations. Generic implementational issues can be found in Carpenter (2003).

2.3 Implementation Details

2.3.1 Data

Ornithological and ecological data used in this study has been obtained from the GAP biodiversity research project of Turkish Society for the Conservation of Nature (DHKD) conducted between 2001 and 2003 (Welch, 2004). Detailed description of observations and data collection method can be found in Kurt (2004) and Welch (2004).

During the field studies, which lasted two years, 1592 points were visited and the ecological variables as well as the breeding success of bird species were recorded. Satellite imagery used in this study was obtained by the Turkish Society for the Conservation of Nature, and consisted of LANDSAT images bands 1-5 and 7, with a resolution of 30×30 m. The characteristics of the satellite images and the properties of the bands used are given in detail in Per (2003) and Kurt (2004).

Independent variables were 6 image bands and 6 environmental variables. Environmental variables were elevation (m), distance to nearest road (m), distance to water (m), vegetation index (categorical), annual relative humidity (%), and annual mean temperature (^{o}C). For all the models considered, the output classes for each data pattern has been assigned either 0 or 1, depending on the occurrence of individuals recorded for each bird species considered here Kurt (2004).

It is important for statistical learning methods to have an input space where the number of data points for each output category (0 and 1, in our case) is approximately balanced to avoid biased estimates (Ripley, 1996). To that end, although there were 1592 data points collected in our data set, the number of data points corresponding to category 1 (i.e., the presence of individuals) were limited (246 - 274, depending on the species), and in order to establish balance, we randomly selected an equal number of data points with output category 0 to the number of points with breeding individuals (category 1) (Hirzel et al., 2002). Thus, the data fed to the models were consisting of 492-548 observations depending on the bird species considered.

The importance of setting aside independent test data, which should not be included during training, to assess the actual performance of a given model has been rigorously emphasized elsewhere (Ripley, 1996; Özesmi and Özesmi, 1999; Hastie et al., 2001; Tan and Beklioglu, 2005). To that end, we randomly split the data sets for each species into two sets with equal number of data points such that the number of data points corresponding to each category were still balanced, and used one set to train the models, while the other to asses the generalizability of the trained models.

2.3.2 Traditional Classification Models

k-NN, LDA, and QDA models were implemented in R-language statistical software (R, 1991). The theoretical considerations and implementation details for these models can be found in Hastie et al. (2001). GLM and

backpropogation models were implemented using NevProp3 software (Goodman, 1996). For backpropogation models, the architecture of the network is optimized step-wise (Özesmi et al., 2005), and the networks with 8, 3 and 10 hidden units were used as final models for *L. senator*, *H. pallida* and *C. brachydactyla*, respectively. Theoretical considerations for feedforward multilayer backpropogation networks can be found in Rumelhart et al. (1986), Bishop (1995) and Ripley (1996), and the implementation details of GLM and backpropogation models for this particular study are given in Kurt (2004).

2.3.3 ARTMAP

ARTMAP was implemented in Matlab version 7 (Mathworks Inc.). All input variables were standardized to zero mean, and units of standard deviation before being fed to all models, but ARTMAP. For ARTMAP, the input variables are standardized such that they are squeezed into a hypercube $C^P \in [0, 1]$, where P is the number of independent features (i.e., dimension of input space). Theoretical considerations for the reason to use this particular standardization for ARTMAP models is beyond the scope of this report, and interested readers are referred to Kosko (1992).

All six models have been trained three times separately for the three bird species, and each trained model is then tested separately on corresponding test sets to asses its generalizability. All models have been trained using bootstrapping and cross-validation to optimize so-called bias-variance trade-off (Hastie et al., 2001).

3 Results and Discussion

3.1 Performance of the Models on Training and Independent Tests

The performances of all five models for all three different bird species on both training and test sets are given in Table 1. For backpropogation models, the performance is given as c-index, which is approximately the area under the ROC curve (Bishop, 1995). For other four models, the performance is given as percent correctly classified. Note that for data sets with perfectly balanced number of data points corresponding to each output category, percent correct measure is equivalent to the c-index measure (Bishop, 1995; Ripley, 1996). Hence, the performance measures of all five methods in our case are compatible. Further note that unlike traditional performance measures such as R^2, a value of 0.5 for percent correct and c-index indicates a performance not better than random.

As evident from Table 1, the performance of neural network models, both ARTMAP and backpropagation, was superior compared to the traditional classification algorithms. For the latter group, especially for LDA and QDA, the data corresponding to *H. pallida* seems to be particularly "difficult", with both models' performance on training set being around random chance level. Among traditional classification models, although k-NN performed better on training set compared to LDA and QDA, it too suffered from low performance on independent test sets.

Backpropogation and GLM method's performance on training sets was considerably better than previous three techniques, and it is especially noteworthy that backpropogation model predicted all of the data points on the training

sets correctly for the data sets of *L. senator* and *C. brachydactyla*. However with respect to training sets, ARTMAP model performed same on these two data sets, and better on set *H. pallida* than backpropogation model. To this end, also note the number of hidden units in backpropogation and the number of formed categories at fuzzy ART module for input vectors (committed nodes) in ARTMAP models (8,3,10 and 2,4,3, respectively). The number of hidden units (or equivalently, of committed nodes) indicate how well the input space is represented as a compressed code in the internal structure of the model (Ripley, 1996; Carpenter et al., 1991a). Considering the fact that the number of compressed representations are equivalent to the degrees of freedom of the model (Bishop, 1995; Ripley, 1996), ARTMAP appears to be more effective in representing the input space, compared to backpropogation method. And it does so without sacrificing the performance on the training set. In addition, the less the degrees of freedom of a model is, the more generalizable it would be (Hastie et al., 2001). The performances of GLM, backpropogation and ARTMAP models on independent test sets also revealed this fact in that the predictive power of ARTMAP was considerably better than the other two, being close to 1 for each of the three independent test cases (Table 1). Thus, at least for the current data set considered, ARTMAP seems to be more robust in characterizing ecological data and predicting the species occurrence, in terms of both training accuracy and generalizability.

3.2 Computational Efficiency

In addition to its superiority in terms of training and test performance, ARTMAP also has the advantage of being computationally much less expensive than feedforward backpropogation networks. For the results presented in this report, backpropogation network required close to 1000 iterations on the complete training set, which approximately took 18 minutes on a P4 1.8GHz PC. Noting that backpropogation models also require architecture as well as free parameter (e.g. learning rate, momentum etc.) optimization, with each model to be trained separately, to achieve best performance, the amount of computational time required grows significantly. On the other hand, fast-learning mode of ARTMAP (Carpenter et al., 1992) enables the network to learn "one-shot deals", that is to learn without iterating the training set. ARTMAP model on fast-learning mode on the same system took approx 10 seconds to train and achieve the performances given in Table 1. In addition, ARTMAP models have only a single external parameter, and consist of two separate self-organizing maps, and as such, they do not require any optimization steps, which renders these family of models to be considerably powerful in terms of computational time required. The non-iterative nature of ARTMAP method also enables new observations to be incorporated to the model as soon as they are obtained, so that the model can be updated with each new observation without any considerable computational effort.

The noticeable performance of ARTMAP model compared to traditional statistical classification techniques, as well as to feedforward multilayer backpropogation method, particularly in terms of generalizability over new data sets suggest that ART-based methods, as presented in this report are potentially robust statistical techniques that can be used instead of already familiar methods. Considering their relatively little computational requirements compared to their closest follower backpropogation models, ART-based models seem to be potential candidates as future predictive models in ecology.

References

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford University Press.

Carpenter, G. (2003). Default ARTMAP. Technical Report CAS/CNS TR-2003-0008, Boston University, Boston, MA.

Carpenter, G. A., Gjaja, M. N., Gopal, S., Markuzon, N., and Woodcock, C. E. (1997). ART and ARTMAP neural networks for applications: Self-organizing learning, recognition and prediction. Technical Report CAS/CNS TR-96-009, Boston University, Boston, MA.

Carpenter, G. A. and Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self organizing pattern recognition architectures. *Neur. Netw.*, 3:129–152.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., and Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Trans. Neur. Netw.*, 3:698–713.

Carpenter, G. A., Grossberg, S., and Reynolds, J. H. (1991a). ARTMAP: supervised real-time learning and classification of nonstationary data by self-organizing neural network. *Neur. Netw.*, 4:565–588.

Carpenter, G. A., Grossberg, S., and Rosen, D. B. (1991b). ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neur. Netw.*, 4:493–504.

Carpenter, G. A., Grossberg, S., and Rosen, D. B. (1991c). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neur. Netw.*, 4:759–771.

Chawla, S., Shekar, S., Wu, W., and Ozesmi, U. (2001). Modelling Spatial Dependencies for Mining Geospatial Data: An Introduction. In Miller, H. and Han, J., editors, *Geographic Data Mining and Knowledge Discovery (GKD).* Taylor and Francis Publishers.

Cohen, M. A. and Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage of competitive networks. *IEEE Trans. Sys. Man. Cybern.*, SMC-13:815–826.

DeValpine, P. (2003). Better inferences from population-dynamics experiments using Monte Carlo state-space likelihood methods. *Ecology*, 84:3064–3077.

Dunk, J. R., Zielinski, W. J., and Preisler, H. K. (2004). Predicting the occurrence of rare mollusks in northern California forests. *Ecol. App.*, 14:713–729.

Fielding, A. H. (1999). *An introduction to machine learning methods*, pages 1–35. Fielding, A. H. (ed.), Machine Learning Methods for Ecological Applications, Kluwer, Dordrecht.

Goodman, P. H. (1996). *NevProp Software, Version 3.* University of Nevada, Reno, NV. http://brain.unr.edu/index.php.

Grossberg, S. (1976a). Adaptive pattern calssification and universal recoding: 1. Parallel development and coding of neural feature detectors. *Biol. Cybern.*, 23:121–134.

Grossberg, S. (1976b). Adaptive pattern calssification and universal recoding: 2. Feedback, expectation, olfaction, illusions. *Biol. Cybern.*, 23:187–202.

Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neur. Netw.*, 1:17–61.

Gutierrez, D., Fernandez, P., Seymour, A. S., and Jordano, D. (2005). Habitat distribution models: are mutualist distributions good predictors of their associates? *Ecol. App.*, 15:3–18.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

Hirzel, A. H., Hausser, J., Chessel, D., and Perrin, N. (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, 83:2027–2036.

Joy, M. K. and Death, R. G. (2003). Assessing biological integrity using freshwater fish and decapod habitat selection functions. *Environmental Management*, 32:747–759.

Kosko, B. (1992). Fuzzy associative memory systems. In Kandel, A., editor, *Fuzzy Expert Systems.* CRC Press, Boca Raton.

Kurt, B. (2004). Güneydogu Anadolu Bölgesinde bazı kuş türlerinin üreme biyolojilerinin yapay nöron agı modeliyle incelenmesi (Study on breeding biologies of some bird species in southeastern Turkey with artificial neural network models). Master's thesis, Istanbul University, Instiute of Natural and Applied Sciences, Istanbul, Turkey. In Turkish.

Lassiter, R. R. and Kearns, D. K. (1977). Phytoplankton population changes and nutrient fluctuations in a simple aquatic ecosystem model. In Middlebrooks, E. J., Flakenborg, D. H., and Maloney, T. E., editors, *Modeling the eutrophication process.* Ann Arbor Science Publishers, 2 edition.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I., and Delacoste, M. (1996). Role of some environmental variables in trout abundance models using neural networks. *Aquat. Living Resour.*, 9:23–29.

Lek, S. and Guegan, J. F. (1999). Artificial neural networks as a tool in ecological modelling - an introduction. *Ecol. Model.*, 120:65–73.

Maron, M. and Lill, A. (2004). Discrimination among potential buloke (*Allocasuarina luehmanii*) feeding trees by the endangered south-eastern red-tailed black-cockatoo (*Calyptorhyncus banksii graptogyne*). *Wildlife Research*, 31:311–317.

Miller, J. (2005). Incorporating spatial dependence in predictive vegetation models: Residual interpolation methods. *Professional Geographer*, 57:169–184.

Moilanen, A. (1999). Patch occupancy models of metapopulation dynamics: Efficient parameter estimation using implicit statistical inference. *Ecology*, 80:1031–1043.

Olden, J. D. and Jackson, D. A. (2000). Illuminating the 'black-box': A randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model.*, 154:135–150.

Özesmi, S. L. and Özesmi, U. (1999). An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecol. Model.*, 116:15–31.

Özesmi, S. L., Özesmi, U., and Tan, C. O. (2005). Methodological Issues in Building, Training, and Testing Artificial Neural Networks. *Ecol. Model.* in press.

Özesmi, U. and Mitsch, W. J. (1997). A spatial habitat model for the marsh-breeding red-winged blackbird (*Agelaius phoeniceus* L.) in coastal Lake Erie wetlands. *Ecol. Model.*, 101:139–152.

Per, E. (2003). Sultansazlıgı ve Tuzla Gölü'nde lojistik regresyon ve yapay sinir agları ile kus habitat modellemesi (Modeling bird habitats in Sultan Marshes and Lake Tuzla using logistic regression and artificial neural networks). Graduation Thesis. Erciyes University, Department of Biology, Kayseri, Turkey. In Turkish.

R (1991). *R-Language Statistical Software*. Free Software Foundation Inc.

Recknagel, F., French, M., Harkonen, P., and Yabunaka, K. (1997). Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Model.*, 96:11–28.

Reyjol, Y., Lim, P., Belaud, A., and Lek, S. (2001). Modelling of microhabitat used by fish in natural and regulated flows in the River Garonne (France). *Ecol. Model.*, 146:131–142.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

Ross, G. G. (1976). Plankton modelling in the Bay of Villefranche. *Journal of Theoretical Biology*, 56:381–399.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 232:533–536.

Scardi, M. (1996). Artificial neural networks as empirical models for estimating phytoplankton production. *Mar. Ecol. Prog. Ser.*, pages 289–299.

Scardi, M. (2001). Advances of neural network modelling of phytoplankton primary production. *Ecol. Model.*, 146:33–45.

Spitz, F. and Lek, S. (1999). Environmental impact prediction using neural network modelling. *J. Applied Ecol.*, 36:317–326.

Tan, C. O. and Beklioglu, M. (2005). Catastrophic-like shifts in shallow Turkish lakes: A modelling approach. *Ecol. Model.*, 183:425–434.

Tan, S. S. and Smeins, F. E. (1996). Predicting grassland community changes with an artificial neural network model. *Ecol. Model.*, 84:91–97.

Underwood, E. C., Klinger, R., and Moore, P. E. (2004). Predicting patterns of non-native plant invasions in Yosemite National Park, California, USA. *Diversity and Distributions*, 10:447–459.

Welch, H. J. (2004). GAP biodiversity research project 2001-2003 Final Report. Turkish Society for the Conservation of Nature (DHKD), Istanbul, Turkey.

Figure Captions

Figure 1: Schematic representation of fuzzy ARTMAP architecture. Input vectors are processed in ART$_a$ module while target categories are processed in ART$_b$ module. Semi-disks represent adaptive weights. For details, see text. (redrawn from Carpenter et al. (1992)).

Figures

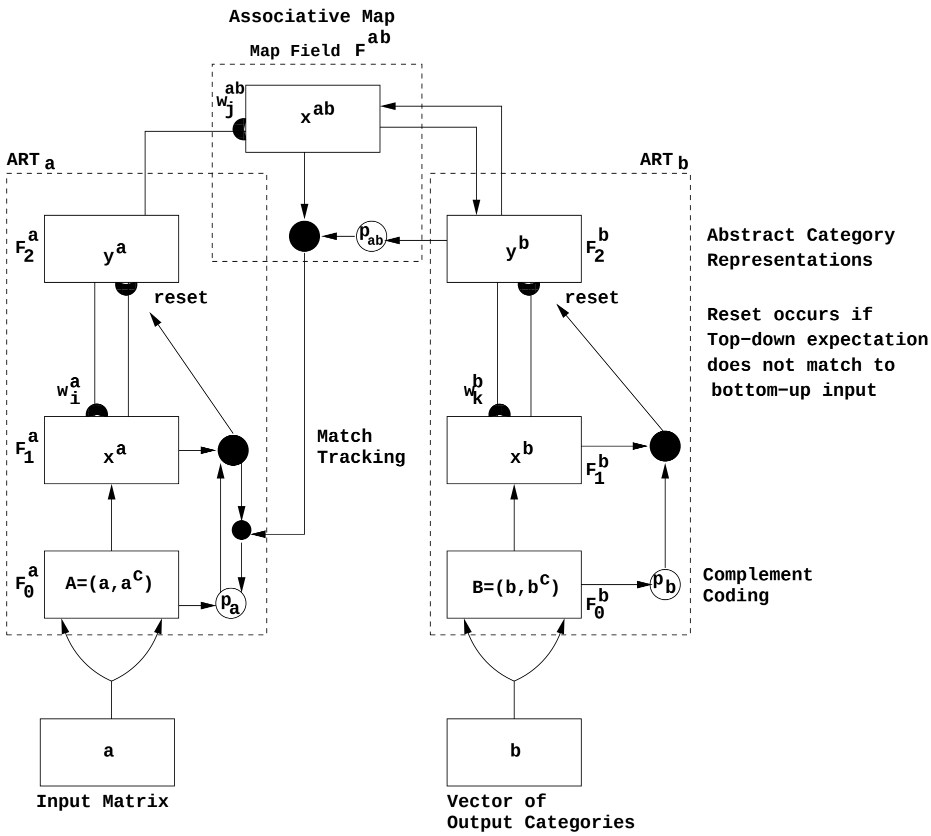

Figure 1:

Tables

Table 1: Performance of the models on training and test sets. *N*: number of data points; *P*: number of input variables; k-NN: k-nearest neighbor; LDA: linear discriminant analysis; QDA: quadratic discriminant analysis; GLM: generalized linear model; BackProp: feedforward multilayer backpropogation network; ARTMAP: adaptive resonance theory based supervised learning. The performance is given as c-index for backpropogation network, and as percent correctly classified for other models (see text).

Set	N	P	k-NN	LDA	QDA	GLM	BackProp	ARTMAP
L. senator(train)	274	12	.828	.781	.799	.859	1.00	1.00
L. senator(test)	273	12	.678	.780	.798	.781	.831	.971
H. pallida(train)	246	12	.866	.488	.496	.759	.874	1.00
H. pallida(test)	245	12	.669	.486	.502	.703	.657	.980
C. brachydactyla(train)	294	12	.847	.646	.701	.855	1.00	1.00
C. brachydactyla(test)	293	12	.765	.648	.703	.769	.809	.962